DOCUMENT RESUME

ED 365 733 TM 020 975

AUTHOR Nweke, Winifred C.
TITLE Enhancing Individual Learning: The Forgotten Role of Educational Assessment.
PUB DATE Nov 93
NOTE 18p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 10-12, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; Correlation; *Educational Assessment; Educational Improvement; Higher Education; *Learning Strategies; *Student Attitudes; Test Construction; *Test Use; Test Validity; *Undergraduate Students
IDENTIFIERS *Paper and Pencil Tests; *Performance Based Evaluation

ABSTRACT

A crucial, but often forgotten, role of educational assessment is to enhance students' learning. This author advocates that an assessment program designed for student learning differs from assessment for accountability in purpose, test format, measurement type, number, and spread of tests, use of test results, and amount of interval between announcement and test administration. This paper illustrates how a combination of traditional paper-and-pencil tests and performance-type assessments has been used to facilitate learning in an undergraduate Test and Measurement class. The sample was 33 students enrolled in the course. The assessment techniques used comprised five pencil-and-paper tests, which accounted for 40% of the course grade, five homework assignments (40%), and a capstone project (20%). The traditional tests and performance assignments were designed to overlap on topics and concepts to reinforce and supplement one another. A low correlation (r=.37) was found between grades on the pencil-and-paper tests and the performance part. Students report that doing the performance assignments engendered and facilitated a better understanding of the material through independent inquiry, problem solving, test construction and validation. Students also indicated that the nonthreatening nature of the projects and homework sustained their hopes of passing the course, contrary to the feedback from the pencil-and-paper tests. One table and one figure are included. (Contains 13 references.) (Author)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

# Enhancing individual learning: The forgotten role of educational assessment

ED 365 733

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

WINIFRED NWEKE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

**Winifred C. Nweke**
**Tuskegee University**
**Tuskegee, Alabama 36088**

**Paper presented at the annual conference of the Mid-South Educational Research Association, November 9-12, 1993. New Orleans.**

TM020975

## Abstract

A crucial, but often forgotten, role of educational assessment is to enhance students' learning. This author advocates that an assessment program designed for student learning differs from assessment for accountability in purpose, test format, measurement type, number and spread of tests, use of test results, and amount of interval between announcement and test administration.

This paper illustrates how a combination of traditional paper-and-pencil tests and performance-type assignments has been used to facilitate learning in an undergraduate Test and Measurement class. The sample was 33 students enrolled in the course. The assessment techniques used comprised five pencil-and-paper tests, which accounted for 40% of the course grade, five homework assignments (40%), and a capstone project (20%). The traditional tests and performance assignments were designed to overlap on topics and concepts to reinforce and supplement one another.

A low correlation (r=.37) was found between grades on the pencil-and-paper tests and the performance part. Students report that doing the performance assignments engendered and facilitated a better understanding of the material through independent inquiry, problem solving, test construction and validation. Students also indicated that the nonthreatening nature of the projects and homework sustained their hope of passing the course, contrary to the feedback from the pencil-and-paper tests.

The purposes of educational assessment could be classified into two broad categories: formative and summative. Formative uses of evaluation include using results of the assessment to improve programs or student learning. In the classroom, this also means using the actual process of assessment or the tasks students perform to effect individual learning (Cronbach, 1984 ; Sax, 1989). Summative uses of evaluation, on the other hand, include accountability, and retention or discontinuation of programs. Tests and measures used for summative purposes have been described as high stakes testing or measurement because of their consequences for policy decisions. Where tests are used for accountability, the program's future, program personnel's credibility or jobs, sometimes, depend on how students perform on standardized tests. This has had the undesirable effect of school systems investing inordinate amounts of money and school time preparing for and taking national examinations. Another effect is what is referred to as Measurement Driven Instruction, (MDI) (Cizek, 1993; Shepard, 1993). In attempt to ensure high scores, schools often teach only such topics or skills that the tests assess. Thus, if the test does not include composition writing, the teachers stop teaching students to write. While tests were initially conceived to serve as thermometers that measure students' level of performance or achievement, under the MDI they become the determinants of curriculum, or agents of change (Porter, 1991; Cizek, 1993).

With all the attention and resources devoted to scoring high on accountability tests, assessment for student learning, the other role of educational assessment is often relegated to the background. Some researchers, such as Frary, Cross, and Weber (1993), even contend

that "the primary purpose of testing in a secondary academic course is and should be for grade determination" rather than student learning as others propose (Sax 1989; Mehrens and Lehmman (1991) and the National Council of Teachers of Mathematics (NCTM, 1992). Teachers often have a list of topics that they are expected to cover each school year. At the same time, teachers are under pressure to drill and coach students so the latter would score high on standardized tests. Consequently, teachers do not have time to help students learn through assessment activities or through the feed back from the many tests that students are subjected to. It has become obvious that the use of educational assessment for individual learning entails different processes from, and cannot effectively compete with, testing for accountability. This realization is evidenced in the call by many educators for a separation of testing for the two purposes (Anrig, 1991, and the National Council of Teachers of Mathematics, 1992).

Assessment designed for student learning differs from that designed for accountability or summative purposes, not only in purpose, but also in its format, type of measurement, the number and spread of measures, and early announcement of the assessment schedule. There is a current shift towards performance tests and away from traditional testing formats. One of the immediate causes of this shift is the performance of American students on international examinations. This has refocused attention on the role and effect of testing in American education. Former President Bush proposed, among other things, not only a national examination but also that performance test format be used to ascertain what students learned. Intuitively, the performance test appears to be more authentic and a better way for students to demonstrate whatever knowledge or skills they

have. Consequently, professional groups, State Departments of Education, and individual teachers and researchers have latched on to this test format and assumed its reliability and validity without any evidence of either (Baron, 1990).

One of the concerns about using performance tests is content sampling and reliability of only one or two such tasks in a test. Some measurement and evaluation specialists have examined, in depth, the issues of reliability, validity, content sampling, and generalizability of performance test results (Mehrens. 1992; Linn, 1993; Yen, 1993; Shavelson, et al 1993). Shavelson et al concluded that students' performance depends, to some extent, on measurement methods used and that these methods tend to elicit different aspects of students' achievement. Their study also shows that large number of tasks using many measurement techniques and over varying occasions is needed to be able to generalize students performance. While the issues of generalizability and reliability may be a major concern for one-shot external examination programs ( Porter, 1991, Linn, 1993), they may not constitute a great problem in classroom testing designed for student learning (Rudman, 1993). Series of performance tests and portfolios spread over the semester can be combined with some pencil and paper tests to obtain multiple measures for student evaluation. Such a combination will eliminate, or at least, reduce the problem of generalizability and reliability of test results that plague one-shot performance tests.

Assessment designed for student learning should be tilted towards criterion-referenced measurement and interpretation. This deemphasizes competition and comparison among students and allows the teacher to help each student learn the material, sometimes, at their own rates and after many trials. That may mean allowing students to redo

assignments, if necessary, after additional clarification.

Assessment designed for student learning should yield multiple measures collected over many occasions. It is agreed in the measurement field, that any obtained score is a function of true score (Xt) and error (Xe). The error in each test score is either in one's favor or against one. There is no way of knowing the magnitude or direction of the error on any one test. However, the sum over many testing situations, is believed to be zero and thus error effect is eliminated by averaging across many measures. Thus, multiple measures from pencil-and-paper tests, performance type tests and assignments, spread over the semester provide a better sampling of occasions and tasks. It is also known that the performance of some students is adversely affected by high levels of debilitating anxiety. Thus, multiple test formats and testing situations, e.g term papers, performance type assignments, portfolios, will provide such students with more varied opportunities to show what they can do. More importantly, multiple measures should emanate from assignments and tests that are arranged in such a way as to reinforce and overlap over concepts, skills and knowledge. Otherwise, multiple measures may just be results of a series of isolated one-shot tests with little or no effect on student learning.

Effective use of feedback is another characteristic of assessment designed for student learning. The feedback is more than merely telling students their grades, or indicating which item is correct or incorrect. It entails a detailed examination of incorrect options to expose incorrect or faulty reasoning, assumptions and mistakes. This type of feedback helps students to improve their test taking skills.

Finally, assessment designed for student learning should not hold any assessment

surprises. In other words, the syllabus should contain, at least, the evaluation schedule for all tests and assignments. This author prefers to see all homework assignments or performance test stimuli included in the syllabus. This way, students can plan ahead regarding when to start to prepare for tests or do assignments.

This study illustrates how a combination of traditional paper-and-pencil tests and performance-type assignments has been used effectively to facilitate learning in an undergraduate Tests and Measurement class. The next section shows the method used.

## Method

### Sample

The sample comprises 33 students with diverse majors enrolled in the Tests and Measurement class in the Spring of 1993 either as a required course or as an elective.

### Procedure and Material

All students are provided with a course syllabus at the beginning of the semester which specifies the behavioral objectives of the course that students are expected to demonstrate or show by the end of the course. It also contains laboratory experiences, a detailed specification of all homework assignments and the project. The syllabus also contained a course outline indicating a week by week plan of work and dates for pencil and paper examinations and due dates for the homework assignments and the project.

Table 1 shows the spread of home-work assignments, the tests, the project and the corresponding chapters or topics which they cover. A major purpose of assessment in this course is to facilitate student learning, not merely for a summary judgment or documentation of whether or not the students passed or failed. Thus, the pencil-and-paper-

tests, homework assignments and the project are designed to overlap over topics and concepts and thus to reinforce one another. Table 1 shows this overlap while Figure 1 shows the same graphically.

The pencil-and-paper tests are combinations of objective and short answer type tests and administered as classroom group tests. The feedback comprises not only telling students their grades but also exploring their wrong choices with them, why they made those choices, why they were wrong and why the correct options were right. The pencil-and-paper tests are not comprehensive. In other words, topics and concepts tested in test 1 are not included in another test except those that are subsumed in later concepts. Thus, the pencil-and-paper tests in effect are as much a one-shot test as the external tests.

The homework assignments are performance type tests that require students to collect data from the school systems, interview teachers, guidance counselors or psychologists regarding their testing practices, test selection and test use. These assignments require the application of concepts and skills from various chapters and thus help students to reinforce and internalize earlier learning. For example, Test 1 covers such topics as the differences and relationships among tests, measurements and evaluation; types of tests and measurements, role and types of objectives in educational evaluation, preparing questions for and grading responses to the essay test. Homework 1, which overlaps with Tests 1 and 2, requires students to obtain from a teacher or professor a set of objectives and a copy of the test that measures its attainment, classify the objectives, classify the test into objective or essay types, classify the test items according to Bloom's Taxonomy, and to make an evaluation of how well the test measures the objectives.

Test 2 covers topics such as writing objective tests, administering, scoring and analyzing classroom achievement tests, and other teacher made evaluation procedures such as performance assessments, sociometry, observation or rating scales. Test 3 covers interpretation of test scores including some descriptive statistics, norms, scores and profiles, reliability, and validity. Homework assignment 2 overlaps with Tests 2 and 3 and requires that students find out, from school psychologists, teachers, or guidance counselor, what types of test they use in their line of work; to classify them in terms of power/speeded, group/individual, self-made/standardized tests; to find out how and why the particular tests were chosen ( e.g for reliability, validity, availability of norms, ease of administration; scoring and interpretation, and test results are interpreted and used; and to evaluate the interviewee's rationale for test selection and interpretation. Finally, to indicate, if they would make similar or different choices if in a similar position.

Homework 3 overlaps with Tests 3 and 4. Test 4 covers the factors affecting measurements of individual, marking and reporting the results of measurements, accountability: testing and evaluation programs and teacher evaluation. Homework 3 requires that students obtain a high school report card, have five parents interpret the same report card, and prepare a report on how well the parents understand what the report card is designed to communicate, what parents would prefer to see or added to the report cards etc. Finally, Homework 4 overlaps with Tests 1 through 5. Test 5 covers topics in standardized evaluation procedures while Homework 4 requires that students use the Oscar Buro's Mental Measurement Year Books and other resources to compare and evaluate two tests that are used for the same or similar purposes, e.g ACT and SAT for college

admissions. In addition to the 4 homework assignments, students construct and validate a 10-20 objective item test to measure the attainment of teacher-specified objectives in any subject area of their choice in the classroom of any teacher in the school system who is willing to cooperate with them. This assignment ties the course together and overlaps with most of the pencil and paper tests.

When each test was returned, some of the students' responses were examined to help students understand the error in reasoning that led to incorrect choice that they made. Though the material was not covered formally in a future test, the feedback was aimed at improving the process rather than the content. For the homework assignments, students asked for and received additional guidance at any stage. Homework assignments that were very badly done were repeated after further clarifications on what was expected. These homework assignments were designed more for students' learning than for determination of grades. Grades were more criterion- than norm-referenced and so it was ethically easier to allow students to redo homework in order to learn the material and consequently improve their scores. Students were not held to their first effort merely for fear of violating some test standardization requirements. Nevertheless, they had only one chance to redo an assignment and also suffered the penalty of not being able to make the maximum possible score. For example, if the redone paper is an A paper, it would be assigned a B considering that it was a second attempt.

## Results and Discussion

Table 2 shows the number of students who would have passed if performance was

based on the five one-shot tests, the performance assignments, and a combination of the pencil and paper tests and the performance tests (homework and project). Only eight students would have passed the class with a grade of either A, B, or C, if success were based on the five one-shot pencil-and-paper tests. On the other hand, as many as 23 would have passed if performance were based only on the performance tests or homework. But when performance is based on a combination of the two formats, 17 students meet and or surpass the 70% pass score.

The correlation between the pencil-and-paper and the performance assignments is low (r=0.37). The percent agreement between pencil-and-paper and performance type tests is 52%. These statistics would be higher if students' scores on their first attempt on the performance assignments are used in the analyses. Unlike policymakers in Connecticut, reported in Baron (1991, p.251), who decided to use performance test results and ignore those from pencil-and-paper, this author combined the results for for the purpose of assigning grades. The number of students who passed under the joint criteria and students' comments indicate some incremental validity of the performance tests. This author agrees with Mehrens (1992) that neither the pencil-and-paper nor the performance test results should be used alone. These formats should provide multiple measures through diverse opportunities for students' overall assessment and more reliable and valid evaluation

Given the low relationship between the pencil and paper tests and the performance tests, it is not advisable to substitute performance test results for pencil and paper tests as reported in Baron (1990). Various researchers, especially in the area of measurement, agree and advocate that assessment program that facilitates student learning should use a

combination of test format. They also advocate that such an assessment program should apply both criterion and norm-referenced measurement and interpretation, and yield multiple measures from numerous tasks spread over many occasions. These characteristics will ensure, or at least improve, reliability, validity and generalizability of decisions.

## REFERENCES

Baron, J. B. (1990). Performance assessment: Blurring the edges of assessment, curriculum, and instruction. In G. Kulm & S. M. Malcom (eds.). Science Assessment in the service of reform, Washington, DC.: American Association for the Advancement of Science, pp. 247-266.

Cizek, G. J. (1993). Rethinking Psychometricians' belief about learning. Educational Researcher 22(4), 4-9.

Cronbach, L. (1984). Essentials of psychological testing. 4th ed. New York: Harper & Row.

Frary, R. B., Cross, L. H. & Weber, L. J. (1993). Testing and grading of academic subjects: Implications for instruction in measurement. Educational Measurement: Issues and Practice, 12(3), 23-30.

Hartel, E. H. (1991). Form and function in assessing science education. In G. Kulm & S. M. Malcom (eds.). Science assessment in the service of reform. Washington, D. C.: Association for the Advancement of Science. PP. 233-245.

Mehrens, W. (1992). Using performance assessment for accountability purposes. Educational Measurement: Issues and Practice, 11(1), 3-9, 20.

Mehrens, W. A. & Lehmman, I. J. (1991). Measurement and evaluation in education and psychology. Chicago: Holt, Rinehart and Winston, Inc.

National Council of Teachers of Mathematics. (1992?). Standard 3: Appropriate assessment methods and uses. NCTM Evaluation Standards. Author.

Rudman, H. C. (1993). National testing or political testing: Is there a difference. Educational Measurement: Issues and Practice, 12(3), 5-9, 30.

Sax, G. (1989). Principles of educational measurement and psychological measurement and evaluation. 3rd. ed. Belmont (CA): Wadsworth Publishing Co.

Shavelson, R. J., Baxter, G. P. and Gao, X. (1993). Sampling Variability of performance assessments. Journal of Educational Measurement 30(3), 215-232.

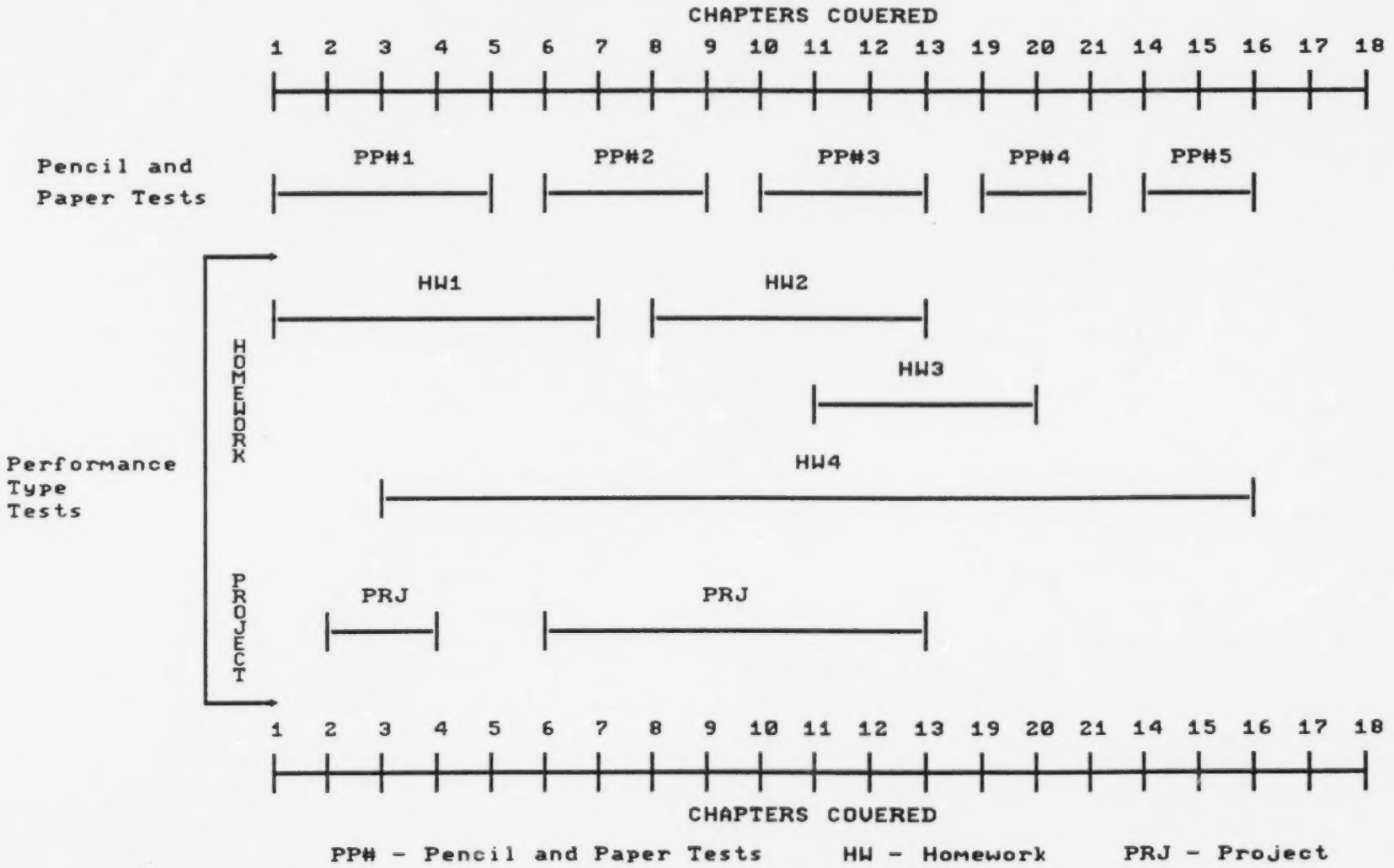
Shepard, L. A. (1993). The place of testing reform in educational reform: A reply to Cizek. Educational Researcher, 22(4), 10-13.

Yen, W. M. Scaling performance assessments: Strategies for managing local item dependence. Journal of Educational Measurement, 30(3), 187-213.

Table 1. Overlapping Distribution of pencil-and-paper tests and performance tests

| Pencil & Paper Tests | | Performance Tests | |
|---|---|---|---|
| Test No. | Chapters Covered | Home-work No. | Chapters |
| I | 1-5 | I | 1-7 |
| II | 6-9 | II | 8-13 |
| III | 10-13 | III | 11-13, 19 & 20 |
| IV | 19-21 + Readings In Teacher Evaluation | IV | 3-16, 19 & 20 |
| V | 14-16 | Project: 2-4, 6-13 | |

Figure: I Graphical Display of the Overlap of Pencil and Paper Tests and Performance Type Tests




PPH - Pencil and Paper Tests HW - Homework PRJ - Project